PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2003-099306
(43)Date of publication of application : 04.04.2003

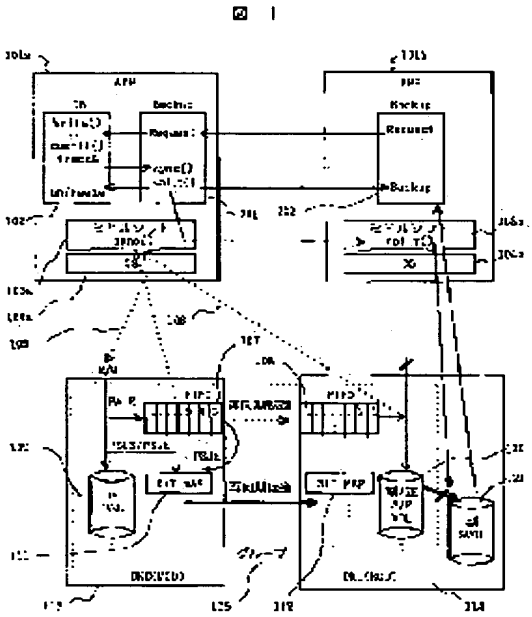
(51)Int.CI. G06F 12/00
G06F 3/06
G06F 12/16

(21)Application number : 2001-290676
(22)Date of filing : 25.09.2001

(71)Applicant : HITACHI LTD
(72)Inventor : URABE KIICHIRO
URATANI IKUO

(54) COMPUTER SYSTEM, AND BACKUP METHOD IN THE COMPUTER SYSTEM

(57)Abstract:
PROBLEM TO BE SOLVED: To solve the problems, when a volume at a remote copy destination created by transfer of data in asynchronous transmission system is backed up, of a remote sub-volume being required to be backed up after the split of the remote volume in order to assure to which time the data is reflected to the remote volume, and also after the split, of the resynchronization of remote volume being required in order to return to the original state and that all the while, the process is carried out without duplication.
SOLUTION: At the start of backup of volume at the destination of asynchronous remote copy, a host backup software submits a sync command right after the freezing of a database (DB) and backs up a remote sub-volume after confirmation that write data, right after the freeze is reflected to the remote sub- volume. For the confirmation of reflection, the newest sequence number of FIFO of a local disk control unit and the newest sequence number of write of a remote disk control device are acquired in the sync command. The sync command compares the newest sequence number, when receiving Sync and the newest sequence number of remote disk control unit, and detects the agreement.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2003－99306
（P2003－99306A）
(43)公開日　平成15年４月４日(2003.4.4)

(51)Int.Cl.⁷	識別記号	ＦＩ	テーマコード(参考)
Ｇ０６Ｆ　12/00	５３１	Ｇ０６Ｆ　12/00	５３１Ｄ　５Ｂ０１８
	５３３		５３３Ｊ　５Ｂ０６５
3/06	３０４	3/06	３０４Ｆ　５Ｂ０８２
12/16	３１０	12/16	３１０Ｊ

審査請求　未請求　請求項の数５　ＯＬ　（全８頁）

(21)出願番号　特願2001－290676(P2001－290676)

(22)出願日　平成13年９月25日(2001.9.25)

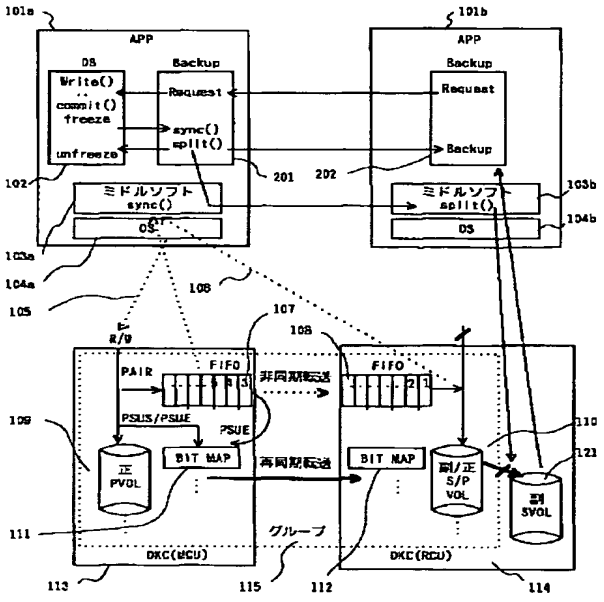
(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地
(72)発明者　占部　喜一郎
神奈川県小田原市中里322番地２号　株式会社日立製作所ＲＡＩＤシステム事業部内
(72)発明者　裏谷　郁夫
神奈川県小田原市中里322番地２号　株式会社日立製作所ＲＡＩＤシステム事業部内
(74)代理人　100068504
弁理士　小川　勝男　（外２名）

最終頁に続く

(54)【発明の名称】　計算機システムおよび計算機システムにおけるバックアップ方法

(57)【要約】
【課題】　非同期転送方式でデータが転送されて作られたリモートコピー先ボリュームをバックアップする場合、リモートボリュームにどの時点までのデータが反映されているかを保証するためリモートボリュームを一旦Splitした後にリモート副ボリュームをバックアップする必要がある。また、Splitした後、元の状態に戻すためにリモートボリュームを再同期する必要がある。この間二重化がなされないで処理が進められる。
【解決手段】　非同期リモートコピー先ボリュームのバックアップ開始時にホストバックアップソフトはデータベース(DB)の凍結直後にSyncコマンドを発行して、凍結直後のWriteデータがリモート副ボリュームに反映されたかを確認した後にリモート副ボリュームをバックアップする。反映の確認のためSyncコマンド内でローカルディスク制御装置のFIFOの最新シーケンス番号とリモートディスク制御装置の書込み最新シーケンス番号を取得する。SyncコマンドはSyncを受取った時点の最新シーケンス番号とリモートディスク制御装置の書込み最新シーケンス番号を比較しその一致を検出する。

図　1



【特許請求の範囲】

【請求項１】ローカルサイトに設けられたローカルディスク装置と、リモートサイトに設けられたリモートディスク装置と、前記ローカルディスク装置に設けられたローカルボリュームに記憶されるデータをシーケンス番号を付与して前記リモートディスクに非同期転送方式で転送する手段と、バックアップ要求に基づいてローカルサイトに設けられたデータベースを凍結する手段と、前記データベースの凍結後前記ローカルディスク装置にあるデータの最新のシーケンス番号と前記リモートディスク装置にある最新のシーケンス番号とを比較しその一致を検出する検出手段とを備えたことを特徴とする計算機システム。

【請求項２】さらに、前記検出手段で一致を検出すると前記リモートディスク装置でのバックアップ開始を指示し前記データベースの凍結を解除する手段を備えたことを特徴とする請求項１記載の計算機システム。

【請求項３】前記リモートディスク装置は前記ローカルボリュームに書き込まれるデータが二重書きされるリモートボリュームと、前記リモートボリュームに書き込まれるデータが二重書きされるリモート副ボリュームを有し、前記検出手段で一致を検出すると前記リモートボリュームと前記リモート副ボリュームをスプリットする指示を出す手段を備えたことを特徴とする請求項１記載の計算機システム。

【請求項４】リモートサイトにあるリモートディスク装置へ自ディスク装置に記憶するデータを非同期転送方式で転送する手段を有するローカルディスク装置と、バックアップ要求を受けると前記ローカルディスクへの書き込み要求を停止せしめる手段と、前記書き込みの停止にあたって最後に発行された書き込み要求のあったデータが前記リモートディスク装置に書き込まれたことを検知する手段とを備えたことを特徴とする計算機システム。

【請求項５】ローカルボリュームを有するローカルディスク装置と、前記ローカルボリュームに書き込まれるデータが二重書きされるリモートボリュームと、前記リモートボリュームに書き込まれるデータが二重書きされるリモート副ボリュームを有するリモートディスク装置とを有する計算機システムにおいて、ローカルボリュームに記憶されるデータにシーケンス番号を付与して前記ローカルディスク装置から前記リモートディスク装置へ非同期転送方式で転送し，バックアップ要求を受けると前記ローカルディスク装置が設けられたローカルサイトのデータベースを凍結し，ローカルディスク装置での書き込み要求データの最新シーケンス番号と前記リモートディスクでの最新のシーケンス番号とを比較し、比較の結果双方のシーケンス番号が一致すると前記リモート副ボリュームをスプリットすることを特徴とする計算機システムにおけるバックアップ方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明はディスク記憶装置間の非同期データ転送によるリモートコピーの作成およびそのリモートコピーのバックアップ作成における現用とリモートのそれぞれのディスク記憶装置の記憶内容の同期制御に関するものである。

【０００２】

【従来の技術】近年のコンピュータシステムは銀行及び証券業務などの基幹業務を大型コンピュータにより一括管理するコンピュータシステムからクライアント・サーバシステムを中心とする分散システムに移行している。このような分散システム環境ではクライアントからの要求を複数のサーバと大型ディスクアレイ装置でデータを処理するHA（高可用性）クラスタ構成が採られている。HAクラスタ構成では地震などの災害に備えて遠隔地にあるデータセンターでデータを二重化する方法が採られている。即ち、あるサイトに設置された大型ディスクアレイ装置のボリューム（ここではローカルボリューム、または正ボリュームという）に格納されるデータを遠隔地に設置された大型ディスクアレイ装置のボリューム（ここではリモートボリューム、または副ボリュームという）にも格納する。このため、通常、大型ディスクアレイ装置間を接続しホスト装置からの書込みデータを大型ディスクアレイ装置間で転送する方法が採られている。遠隔地へデータを送りボリュームのコピーを作成することをリモートコピーという。

【０００３】大型ディスクアレイ装置間で書き込みデータを転送する方法は大別すると同期転送方式と非同期転送方式の２種類がある。同期転送方式はデータを転送する度にそのデータの受領を示す情報を受け取る方式であり，比較的近距離でのデータの転送に適する。一方，非同期転送方式はデータを一方的に応答を受けないまま転送するもので、遠距離でのデータの転送に適する。リモートコピーには一般に非同期転送方式が採られている。

【０００４】図４は従来技術の全体構成を示したシステムブロック図であり、一般的なHAクラスタの全体構成図である。図で符号の最後にａが付されているものは正側の装置またはソフトウェアを示し，ｂが付されているものは副側の装置またはソフトウェアを示す。１６ｂの転送路に交差して付された線は通常時はホスト装置１ｂからの書き込みが禁止されていることを示す。

【０００５】HAクラスタを構成するホスト装置１ａ，１ｂはデータベース等のAPP２ａ，２ｂ、ミドルソフト３ａ，３ｂ、OS４ａ，４ｂでそれぞれ構成される。ミドルソフト３はリモート側（副サイト）では正サイト障害での副サイト運用時に副サイトのホスト装置１ｂから大型ディスクアレイ装置１３ｂへのデータの書込み禁止状態を解除し、正サイト側では初期ペア状態の生成（ペア状態とは正ボリュームと副ボリュームが存在し、その双方にデータが二重に記憶される状態をいう。初期ペア状態

の生成とは副サイトに二重書きされるべきボリュームを設定することをいう。）、ペア状態のサスペンド（二重書きを停止し、正サイトのボリュームにだけ書き込みが行なわれるようにすること）等のペア制御指令を大型ディスクアレイ装置１３に送るコマンド制御ソフトである。

【０００６】ホスト装置１ａからのIO要求はIOI/F（入出力インタフェース）２３ａを介して大型ディスクアレイ装置１３ａに送られる。大型ディスクアレイ装置１３ａはHOST I/F制御１７ａでIO要求を受けキャッシュ２１ａにWriteデータを書込む。このキャッシュ２１ａに書かれたデータはドライブ制御２２ａによって実際の物理ディスク９ａに書込まれる。通常、大型ディスクアレイ装置１３ａ，ｂは複数の物理ディスク９ａ，ｂをRAID１、RAID５などの構成とすることによって保護し物理ディスクの障害に備えている。

【０００７】大型ディスクアレイ装置間でのデータの二重化では、キャッシュ２１ａに書かれたデータはリモートI/F制御１８ａを介してリモート大型ディスクアレイ装置１３ｂに送られ、リモート大型ディスクアレイ装置１３ｂ上のキャッシュ２１ｂに書込まれる。リモート側でのこの後の動作は前述の物理ディスク９ａへの書込みと同じである。この二重化された状態ではリモート側のHOST I/F制御１７ｂはリモートホスト１ｂからの書込みを禁止する。同期転送方式ではホスト装置１ａからのIO要求に対して、IO要求毎に大型ディスクアレイ装置１３ｂが大型ディスクアレイ装置１３ａを介して前述した手順でリモート大型ディスクアレイ装置１３ｂ上のキャッシュ２１ｂに書かれたことを確認して応答を返す。つまり、ホスト装置１ａからのIO完了はリモート側に書かれたことが保証される。この同期転送方式はローカルとリモート間でデータ伝送の伝播遅延が少ない比較的近距離（100Km以内）に適しているが、公衆回線網を使用する遠距離転送には適さない。一方、遠距離転送に向いている非同期転送方式ではホスト装置１ａからのIOはキャッシュ２１ａに書かれた時点でIO完了を応答する。リモート側へ未転送のデータはBITMAP１１によって差分管理され、この差分データをホスト装置１からのIOに同期しないで非同期にリモート側に転送される。差分管理について更に説明する。差分とは正側のキャッシュ２１ａには書き込まれているが副側のキャッシュ２１ｂには書き込まれていないデータである。その差分データが正ボリュームのどの位置のデータであるかを示すのがBITMAP（ビットマップ）１１ａである。

【０００８】この非同期にリモート側に転送する方法は差分データをホストIOの順序性無しで転送する方法と、ホストIOの順序性を維持して転送する方法があるが何れの方法でもホストIOのリモート側への同期完了確認（リモート側のキャッシュ２１ｂにローカル側のキャッシュ２１ａに書き込まれているものが同じものに達したことを確認すること）が困難であった。このため、この非同期転送方式によるリモートコピー先ボリュームをバックアップする（ローカル側のボリュームが使えない状態になったとき、更にリモート側のボリュームまでが使えない状態になることがあることに備えリモート側のボリュームのバックアップコピーを残す）場合、リモートボリュームのローカルボリュームとのペア状態を一旦サスペンドし、転送中でありリモート側のキャッシュに書き込まれているかが不確定なデータをなくすことにより、どの時点までリモートボリュームがローカルボリュームと一致したデータを記憶しているかを確定させることでデーター致性を保証する必要があった。ここでは、ローカルボリュームとリモートボリュームがペア状態をサスペンドすることをスプリット(Split)とも言う。このペア状態をサスペンドしている期間はローカルボリュームにのみホスト１ａの書き込みデータが記憶されリモートボリュームにはそのデータが記憶されないためシステムの安全性上好ましくない期間である。

【０００９】このように、バックアップソフトはリモートボリュームのローカルボリュームとのペア状態をサスペンド（Split）した後にリモートボリュームをバックアップする。つまり、リモートバックアップのためにリモートボリュームをSplitする必要があった。また、Splitし、リモートボリュームのバックアップファイルを作成した後、元の状態に戻す（ローカルボリュームとリモートボリュームがペア状態になり、サスペンドの期間にローカルボリュームにだけ書き込まれていたデータをリモートボリュームにも書き込むための処理を行なう、具体的には差分データをビットマップに従いリモートボリュームへ書き込むための処理を行なう）即ち,リモートボリュームを再同期（Resync）する必要がある。災害リカバリの観点からみると、ボリュームバックアップのためのペア状態のサスペンド（Split）及び再同期（Resync）は二重化状態を一旦解除しているので、ボリュームを二重化することによって実現されていたHA構成の信頼性を低下することになる。

【００１０】

【発明が解決しようとする課題】上述したように、非同期転送方式でリモートボリュームを作成し、さらにそのリモートボリュームのバックアップファイルを作成する場合、従来の方式では、そのバックアップファイルがどの時点での状態かを特定するために、リモートボリュームのローカルボリュームとのペア状態を一旦サスペンドする必要がある。このように、上述のようなペア状態を維持したままでどの時点でのバックアップファイルを作成したかを確定する方法を備えていない。これは即ち，バックアップ開始時に、この時点での最後のWrite要求データがリモートサイト（副サイト）のキャッシュ２１にWriteされたかを確認する方法がないということである。本発明の目的はリモートボリュームのペア状態を維

持したままリモートボリュームのバックアップ開始時点での正サイトのディスクアレイ装置と副サイトのディスクアレイ装置との同期確定を可能にしバックアップを行うことにある。

【００１１】

【課題を解決するための手段】本発明は、上記課題を達成するために、正側ホストのBackupコーディネータは非同期リモートコピー先ボリュームのバックアップ開始時にミドルソフトと連携しデータベース(DB)を凍結し、Write要求も一時的に止める。その直後にミドルソフトからSyncコマンドを発行する。ホストからのWrite要求データには正側大型ディスクアレイ装置でシーケンス番号が付与され、副側大型ディスクアレイ装置に転送される。副側ではシーケンス番号によって受け取ったデータの順序性を保つと共に、副側大型ディスクアレイ装置のキャッシュに書き込まれたデータの中の最新のシーケンス番号を正側大型ディスクアレイ装置に返す。

【００１２】SyncコマンドはSyncを受取った時点の正側大型ディスクアレイ装置のWrite要求データに付与された最新シーケンス番号と返送されてきた副側大型ディスクアレイ装置内のキャッシュに書込まれた最新シーケンス番号を比較し、一致したときに副サイトのキャッシュへのWrite確定を検知しBackupコーディネータに同期完了を報告する。

【００１３】このSync完了によってBackupコーディネータはDB凍結直後のWriteデータが副側ボリュームに反映されたかを確認でき、Backupリクエスタに副側ボリュームのバックアップの開始を通知するとともにDBの凍結を解除しWrite要求の再開を通知する。

【００１４】

【発明の実施の形態】以下、本発明の一実施例を図１～図３により詳細に説明する。図１は本発明の計算機システムの全体構成を示す論理ブロック図である。図１の基本的なハードウエア構成は図４のブロック図と同じであり図１は本発明の論理的なブロック図を示す。

【００１５】ホスト装置１０１ａはアプリケーションソフトであるデータベース(DB)１０２、Backupコーディネータ２０１、大型ディスクアレイ装置１１３内のボリュームの初期ペア状態の生成、ペア状態のサスペンド等のペア制御を実行するミドルソフト１０３ａ、及びOS１０４ａを有する。一方,ホスト装置１０１ｂはBackupリクエスタ２０２を有する。大型ディスクアレイ装置１１３、１１４はローカルサイトとリモートサイトに置かれ互いに光ファイバや広域回線網などで接続される。大型ディスクアレイ装置１１３，１１４内のボリュームはPVOL１０９とS/PVOL１１０とSVOL１２１とがある。ホスト装置１０１ａからのデータはPVOL１０９からS/PVOL１１０にコピーされ二重化される。BITMAP１１１、１１２はPVOL１０９とS/PVOL１１０間のデータ差分管理テーブルでありPVOL１０９とS/PVOL１１０の全データブロックを数１０KB単位でビットマップ化したものである。通常、ペア状態（二重化状態）がサスペンド(PSUS)した状態でのホストからのデータはPVOL１０９とS/PVOL１１０の不一致としてこのBITMAP１１１、１１２によって差分管理される。

【００１６】FIFO１０７，１０８は大型ディスクアレイ装置１１３、１１４間の非同期転送用のバッファでありペア状態の時に使用される。ホスト１０１ａからのIOの書込みデータは大型ディスクアレイ装置１１３にある正のボリュームであるPVOL１０９のキャッシュに置かれ物理ディスクに書込まれると同時にそのIO単位にシーケンス番号を付加し、一旦FIFO１０７にキューイングされる。このシーケンス番号が付加されたデータは非同期に大型ディスクアレイ装置１１４に転送され、大型ディスクアレイ装置１１４内でシーケンス番号順にFIFO１０８にキューイングされる。このFIFO１０８にキューイングされたデータはシーケンス番号順にS/PVOL１１０のキャッシュに置かれ物理ディスクに書込まれる。

【００１７】大型ディスクアレイ装置１１３、１１４間の転送障害によって非同期転送が出来ない場合、FIFO１０７にキューイングされた未転送データはBITMAP１１６に差分データとして管理され二重化を障害サスペンド(PSUE)状態にする。このように大型ディスクアレイ装置１１３は状態情報を持っている。ホスト１０１ａのミドルソフト１０３ａは大型ディスクアレイ装置１１３の状態をチェックし状態がペア状態であれば二重化状態であると認識し非同期転送中であることを知る。PSUS,PSUE状態（PSUSはsplit指示が出て二重化が解かれた状態、PSUEは障害発生により二重化が解かれた状態を意味する）であれば二重化がサスペンド状態であると認識する。ホスト１０１ａのミドルソフト１０３ａは大型ディスクアレイ装置１１３の状態とFIFO１０７，１０８のキューのPVOL１０９とSVOL１１０のデータシーケンス番号を比較することで直前のWrite要求のデータがリモートサイトに同期したかを確認することが可能である。最新のシーケンス番号が一致していれば同期していると認識できる。

【００１８】ここで、ホスト１０１ａのデータベース(DB)１０２とBackupコーディネータ２０１、ミドルソフト１０３ａと大型ディスクアレイ装置１１３及びリモートホスト１０１ｂ内のBackupリクエスタ２０２間の連携によってホスト１０１ａ上のデータベース(DB)１０２がどの様にリモートバックアップ(リモートサイトでバックアップファイルが作られること)されるかを説明する。

【００１９】まず、リモートホスト１０１ｂ内のBackupリクエスタ２０２はバックアップ開始時にバックアップ要求をホスト１０１ａのBackupコーディネータ２０１に送る。Backupコーディネータ２０１はこの要求を受けるとこの時点のデータベース(DB)１０２を凍結するための要求を送る。データベース(DB)１０２は全てのトランザ

クションをCommitして（トランザクションのデータは逐一データベースに書き込まれないで一時保管されている。これをディスクに書き込む。）データベースを一時的に凍結する。通常はこの凍結でバックアップを開始できるが、大型ディスクアレイ装置間でデータが二重化構成の場合、この凍結直後のデータがリモートボリュームのS/P(副/正)VOL１１０に反映されたか確認する必要がある。

【００２０】この同期確認のためにSyncコマンドを発行する。ここで、Syncコマンドはホスト装置上で動作するライブラリ及びホストコマンドであり、ソフトウエア製品としてCDROM、フロッピー（登録商標）ディスク等で提供される。このSyncコマンドはミドルソフト１０３ａによって提供されパラメータとしてグループ１１５のグループ名（一連の通番で管理されるボリューム群、この中でデータの順序性などを保証する）groupと最大の同期完了待ち時間を指定するtimeoutによって定義される。ミドルソフト１０３ａから発行されたSyncコマンドはペア状態をチェックしペア状態であればFIFO１０７のPVOLの最新のシーケンス番号を取得しこのPVOLの最新シーケンス番号をデータベース(DB)１０２凍結直後のシーケンス番号として保持する。

【００２１】次に大型ディスクアレイ装置１１３を介してリモート側のSVOLの書込みシーケンス番号（リモート側のキャッシュに書き込まれているデータのシーケンス番号）を取得し「PVOLのシーケンス番号≦SVOLの書込みシーケンス番号」が成立するまでSVOLの書込みシーケンス番号を繰返しテストしこのコマンド内で待つ。条件が成立すると同期完了の応答としてこのSyncコマンドは呼び出しもとであるBackupコーディネータ２０１に戻る。Backupコーディネータ２０１はこのSyncコマンドが完了したことでローカルボリュームとリモートボリュームが同期されたとみなす。

【００２２】次にBackupコーディネータ２０１はS/P(副/正)VOL１１０のバックアップボリュームであるS(副)VOL１２１をスプリットするためにミドルソフト１０３ａからsplitコマンドを発行する。このsplitコマンドはミドルソフト間で通信されリモートサイトのホスト１０１ｂを介して大型ディスクアレイ装置(RCU)１１４にS(副)VOL１２１のスプリット指示を発行する。このスプリット指示によってS/P(副/正)VOL１１０のデータがS(副)VOL１２１に反映される。

【００２３】Backupコーディネータ２０１はsplitコマンド発行後、Backupリクエスタ２０２にバックアップ開始を許可し、スプリットされたSVOL１２１のバックアップファイルを作成せしめる。更にデータベース(DB)１０２に凍結を解除し再開を指示する。

【００２４】この許可を受けて、Backupリクエスタはローカルボリュームとリモートボリューム間のペア状態を維持したままリモートボリュームの副ボリュームであるS(副)VOL１２１のバックアップを実行することが可能である。

【００２５】バックアップの作成について、まとめて説明する。ローカルサイトにあるローカルボリューム（現用あるいは正ともいう）とリモートサイトにあるリモートボリューム(副ともいう)との間で非同期転送方式で二重書きする構成とする。更に、リモートサイトではリモートボリューム（ローカルボリュームを正とするなら副に当たる）を正ボリュームとする副ボリュームに二重書きをしている。ローカルボリュームからリモートボリュームにデータが転送されるときにはデータにシーケンス番号が付与されている。バックアップ時にはデータベースを一時凍結する。そして、ローカルサイトでの最新のシーケンス番号とリモートサイトでの最新のシーケンス番号とを比較し、これらが一致するとローカルボリュームとリモートボリュームの間のペア状態を維持したまま、リモートサイトにあるリモートボリュームとその副ボリュームとの間をスプリットする。その後、バックアップ開始を許可し、データベースの凍結を解除し処理の再開を指示する。スプリットされた副ボリュームはバックアップ用ボリューム(用途によってはチェックポイントファイル)であり、これによって,図示されていないバックアップファイルが作成され,これがテープ媒体に格納される。一連のバックアップ処理が終わるとリモートボリューム（リモートサイトでの正）とスプリットされていた副ボリュームを再び同期させる。処理が実行されている間ではローカルボリュームとリモートボリュームのペア状態がサスペンドされることがなく、高信頼性は保たれる。データベースが凍結されている時間は短いものであり、凍結解除後処理が再開されている間並行してバックアップ処理が実行され得る。

【００２６】図２は本発明の一実施例の全体制御フローを示した図である。以下制御フローに基づいて詳細に説明する。まず、制御フローはホスト１０１ａのアプリケーションソフトであるデータベース(DB)１０２と、バックアップシーケンスを制御するBackupコーディネータ２０１と、バックアップを実行するBackupリクエスタ２０２の個々の制御を示す。

【００２７】リモートサイト上のバックアップアプリケーションであるBackupリクエスタ２０２のBackup要求ステップ５１ではバックアップ指示を受けるとまず、ホスト装置１０１ａ上のBackupコーディネータ２０１にBackup要求を送りSplit待ちの状態５２に遷移する。通常、ホスト装置１０１ａ上のBackupコーディネータ２０１はIdle状態５４でありBackup要求の待機状態にある。ホスト装置１０１ａ上のBackupコーディネータ２０１はBackup要求を受けるとDB凍結要求ステップ５５に移りこの時点でデータベース(DB)１０２を凍結するためにDB凍結要求をデータベース(DB)１０２に送りDB凍結待ち状態５６に遷移する。データベース(DB)１０２は通常、オンライ

ントランザクション等の処理を実行しているがこの要求を受けると全てのトランザクションをコミットするためにcommitステップ６１を実行する。更にデータベースを一時的に凍結するためにDB凍結実行ステップ６２に移行してデータベースを凍結し、DB凍結応答ステップ６３に移行してBackupコーディネータ２０１に対してDB凍結応答を送る。DB凍結待ち状態５６のBackupコーディネータ２０１はこの応答を受けてSyncステップ５７に移行しSyncコマンドをミドルソフト３に発行する。ミドルソフト３でのSyncコマンドの動作は後述の図３で説明する。

【００２８】Backupコーディネータ２０１はSyncステップ５７が完了した時点で正ボリューム１０９とリモートボリュームのS/P(副/正)VOL１１０とが同期されたと判断してSplitステップ５８に移行しSplitコマンドを実行する。このSplitコマンドはミドルソフト１０３ａ、１０３ｂ間で通信され、リモートサイトのホスト１０１ｂを介して大型ディスクアレイ装置(RCU)１１４に発行され、S/P(副/正)VOL１１０のデータがS(副)VOL１２１に反映されてS/P(副/正)VOL１１０とS(副)VOL１２１との関係がスプリットされる。Backupコーディネータ２０１はこのSplitステップ５８が完了後Split応答ステップ５９に移行し、Split応答をBackupリクエスタ２０２に返す。

【００２９】Split待ち状態５２で待機しているBackupリクエスタ２０２はBackup要求の応答として受け取りS(副)VOL１２１のバックアップを実行する。更に、Backupコーディネータ２０１はこのバックアップの実行の開始と同時にデータベースを再開するためにDB解凍要求ステップ６０に移行しDB解凍要求をデータベース(DB)１０２に送る。データベース(DB)１０２はこの要求を受けてDB再開ステップ６４を実行し凍結状態にあったデータベースを解凍し再開する。

【００３０】図３はSyncの制御フローを示した図である。以下制御フローに基づいて詳細に説明する。まず、制御フローはホスト１０１ａのアプリケーションソフトであるBackupコーディネータ２０１とペア制御及びSyncコマンドを実行するミドルソフト１０３と、大型ディスクアレイ装置１１３の個々の制御を示す。

【００３１】Buckupコーディネータ２０１はデータベースへのコミット完了直後にsync(group,timeout)３１を発行する。syncコマンドの第1引数であるgroupは前述したグループ名を指定する。第２引数であるtimeoutは最大の同期完了待ち時間を指定する。ミドルソフト１０３はsync(group,timeout)３１を実行する。sync(group,timeout)３１はこのコマンド内でまず大型ディスクアレイ装置１１３内のPVOL１０９のペア状態を調べるためにPVOL状態取得コマンド３２を大型ディスクアレイ装置１１３に発行する。

【００３２】大型ディスクアレイ装置１１３はこのコマンド応答としてPVOL状態応答３９をしPVOL１０９のペア状態を返す。syncコマンドはPVOL状態のチェック３３によって状態がPAIR以外(PSUS,PSUE)であれば二重化はサスペンドであるとして同期失敗を返す。状態がPAIRであれば二重化状態であり書き込みデータはFIFO１０７にキューイングされているので最新のPVOLシーケンス番号（ローカルサイトにあるデータに付されているシーケンス番号を便宜上PVOLシーケンス番号と呼び、それがリモートサイトに転送されるとSVOLシーケンス番号と呼ぶこととする）を調べるためにPVOLシーケンス番号取得コマンド３４を大型ディスクアレイ装置１１３に発行する。大型ディスクアレイ装置１１３はこのコマンド応答としてFIFO１０７上にキューイングされているPVOLの最新のシーケンス番号の応答４０をし最新のPVOLシーケンス番号を返す。こうして取得されたPVOLシーケンス番号は同期確認の間保持されSVOLシーケンス番号との比較に使用される。

【００３３】次に、PVOLシーケンス番号とリモートサイトのSVOLシーケンス番号の比較のためにSVOLシーケンス番号取得コマンド３５を大型ディスクアレイ装置１１３に発行する。大型ディスクアレイ装置１１３はリモートサイトの大型ディスクアレイ装置１１４からS/PVOL１１０に書き込まれた最新のシーケンス番号を取得し、大型ディスクアレイ装置１１３はこのコマンド応答としてSVOL書込みシーケンス応答４１をし最新のSVOLシーケンス番号を返す。次のステップ３６で保持していたPVOLシーケンス番号とSVOLシーケンス番号を比較し、PVOLシーケンス番号 ≦ SVOLシーケンス番号であれば当該PVOLシーケンス番号はSVOL側に書き込み済みであるとして同期完了を返す。PVOLシーケンス番号 ＞ SVOLシーケンス番号であれば同期が完了していないので次の待ちステップに進みタイムアウトのチェック３７を行い指定timeout値を超えていれば同期完了タイムアウトとして同期失敗を返す。指定timeout値を超えていなければ一定時間WAIT３８しステップ３５から同期が完了するまで繰り返しステップ３６でPVOLシーケンス番号≦ SVOLシーケンス番号の条件が成立した時点で同期完了し、Syncコマンド呼び出し元のBackupコーディネータ２０１に制御が戻る。Backupコーディネータ２０１はSyncコマンドの制御が戻った時点で戻り値をチェックし同期完了を確認する。

【００３４】以上説明したように本実施例においては次のような効果が得られる。

(1)ホスト装置のバックアップソフトはボリュームバックアップのタイミングでSyncコマンドとsplitコマンドを発行することでローカルボリュームとそのローカルボリュームのデータの二重化のためのリモートボリュームのペア状態のサスペンド（Split）なしにリモートボリュームをバックアップすることが可能である。

【００３５】(2)リモートサイトでのバックアップのためにローカルボリュームとリモートボリュームとの間のペア状態をサスペンド（Split）する必要がないので、

災害時においてのリカバリの観点からリモートミラー(リモートサイトにミラーボリュームを作ること)を採用するHA構成の信頼性を維持することが可能である。

【００３６】(3)ローカルサイトからリモートサイトへボリュームの二重化のために非同期転送方式に従ってデータを転送しても、バックアップ時には同期転送方式でデータを転送したときと同様にローカルサイトとリモートサイトでのデータの格納状態の同期が確認出来る。

【００３７】（４）非同期転送方式でリモートサイトに二重化ボリュームを作成する場合でもリモートボリュームにそれに対する副のボリュームを設け、副のボリュームのデータがローカルボリュームと同期していることを確認することにより、ローカルボリュームとリモートボリュームの間の二重書きの関係を解除することなくリモートボリュームのバックアップをとることが出来る。

【００３８】

【発明の効果】本発明によればローカルサイトからリモートサイトへボリュームの二重化のために非同期転送方式に従ってデータを転送しても、バックアップ時にはローカルサイトとリモートサイトでのデータの格納状態の同期が確認出来る。

【図面の簡単な説明】

【図１】本発明の一実施例における計算機システムの論理的なブロック図。

【図２】本発明の一実施例における全体制御フローを図１の論理的なブロック図をもとに示した図。

【図３】本発明の一実施例におけるsyncコマンドの動作を説明するための全体制御フロー図。

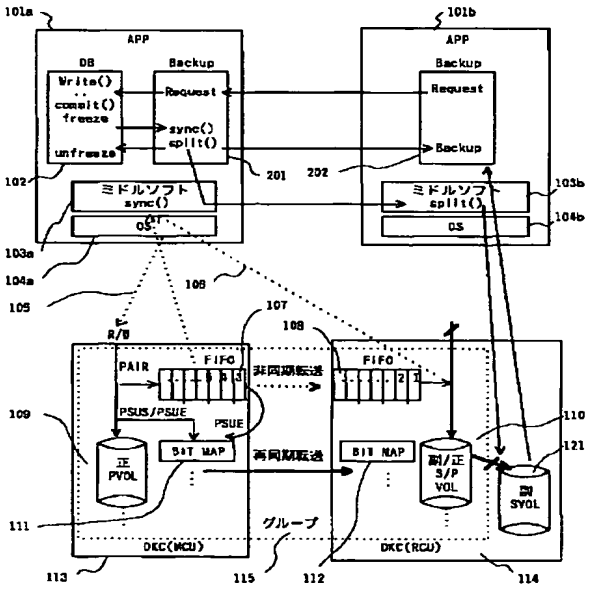
【図４】従来技術の計算機システムの全体構成を示したシステムブロック図であり、一般的なHAクラスタの全体構成図。

【符号の説明】

１０１...ホスト装置，１０２...データベース(DB)，２０１...Backupコーディネータ，２０２...Backupリクエスタ，１０３...ミドルソフト，１０４...オペレーティングシステム(OS)，１０５...PVOL最新シーケンス番号，１０６...SVOL書込みシーケンス番号，１０７...FIFO(PVOL)，１０８...FIFO(SVOL)，１０９...P(正)VOL，１１０...S/P(副/正)VOL，１２１...S(副)VOL，１１１...BITMAP(PVOL)，１１２...BITMAP(SVOL)，１１３...大型ディスクアレイ装置(MCU)，１１４...大型ディスクアレイ装置(RCU)，１１５...グループ

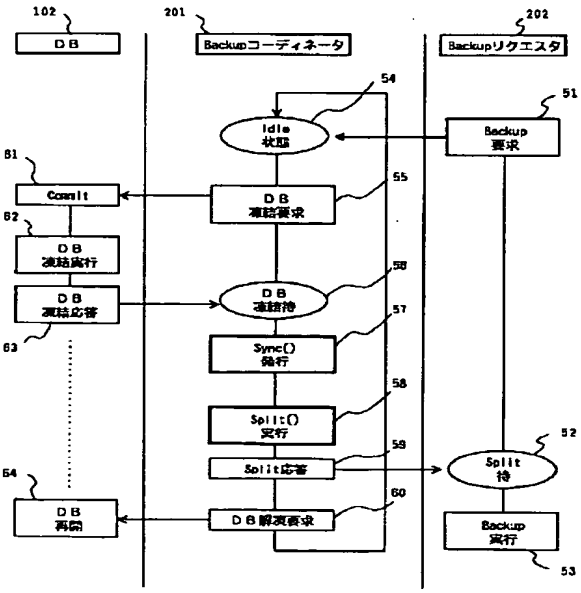
【図１】

図 1



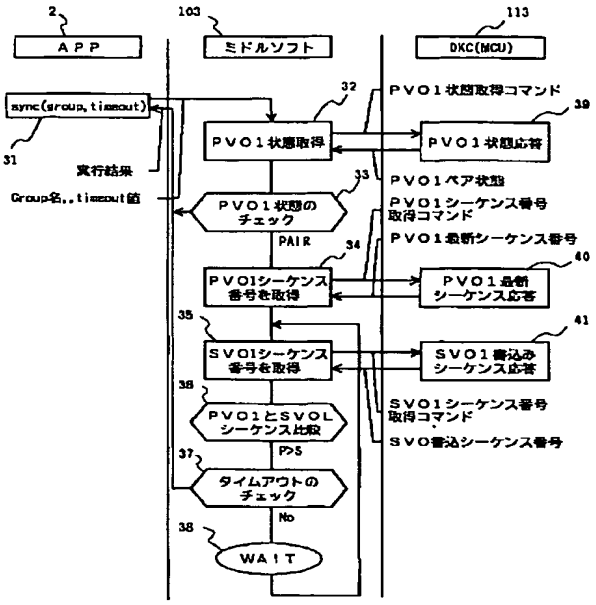
【図２】

図 2



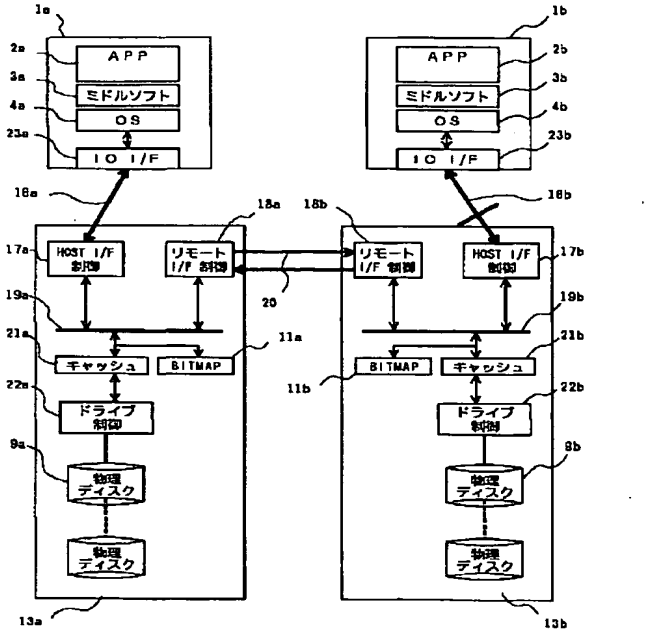
【図３】

図 3



【図４】

図 4



フロントページの続き

Ｆターム(参考)　5B018 GA04 HA04 KA03 MA12　　30
5B065 BA01 EA31 EA35 ZA01 ZA15
5B082 DE04 DE07 GA04 GB02 GB06
HA03